Consensus Matching Pursuit of Multi-Trial Biosignals, with Application to Brain Signals

Christian G. Bénar, Théodore Papadopoulo, Bruno Torrésani, Maureen Clerc

12345

Consensus Matching Pursuit of Multi-Trial Biosignals, with Application to Brain Signals

Christian Bénar,Théodore Papadopoulo,Bruno Torrésani,and Maureen Clerc

Abstract

Time-frequency representations are commonly used to analyze the oscillatory nature of bioelectromagnetic signals. There is a growing interest in sparse representations, where the data is described using few components. In this study, we adapt the Matching Pursuit of Mallat and Zhang for biosignals consisting of a series of variations around a similar pattern, with emphasis on multi-trial datasets encountered in MEG and EEG.

The general principle of Matching Pursuit (MP) is to iteratively subtract from the signal its projection on the atom selected from a dictionary. The originality of our method is to select each atom using a voting technique that is robust to variability, and to subtract it by adapting the parameters to each trial. Because it is designed to handle inter-trial variability using a voting technique, the method is called Consensus Matching Pursuit (CMP).

The method is validated on both simplified and realistic simulations, and on two real datasets (intracerebral EEG and scalp EEG). We also compare our method to two other multi-trial MP algorithms: Multivariate MP (MMP) and Induced activity MP (IMP). CMP is shown to be able to sparsely reveal the structure present in the data, and to be robust to variability (jitter) across trials.

I. Introduction

Bioelectromagnetic signals such as the electroencephalogram (EEG), magnetoencephalogram (MEG) or electromyogram (EMG) can only be recorded with very high-gain amplifiers. In spite of amplification, the contribution of the activity of interest to the recorded signal is often very

C. Bénar is with INSERM U751, Université de la Méditerranée, Marseille, France e-mail: christian.benar@univmed.fr.

T. Papadopoulo is with Odyssée Project Team, INRIA, France.

M. Clerc are with Odyssée Project Team, INRIA and Universit Paris Est, Ecole des Ponts ParisTech, CERTIS, France.

B. Torrésani is with LATP, CMI, University of Provence, Marseille, France.

Manuscript received xx; revised xx.

weak, because of noise contamination. The origin of such noise is twofold: the non-biological 'environmental' noise, which can be reduced by appropriately designing the recording conditions, and the biological 'background' noise - impossible to prevent - which can loosely be defined as the part of the signal which is unrelated to the study at hand. In order to extract relevant information from these recordings, an experiment is generally repeated over many trials, and the data is averaged across those trials in order to improve the signal to noise ratio (SNR).

The underlying assumptions of the traditional averaging methods are that (i) the signal of interest is identical across trials and (ii) the noise is additive, stationary, and zero-mean. Both assumptions are questionable. In this paper, the reliance on the first assumption is relaxed. Indeed, even within the same recording session and for the same subject, it is recognized that the signals may present strong inter-trial variability in phase, latency or amplitude (e.g., [1], [2]). Averaging data across trials without accounting for this variability distorts the waveforms and may result in a loss of information. Furthermore, assuming the signal of interest to be identical across trials, when in fact it is not, leaves residual signal when subtracting an estimated signal waveform. This results in an estimate of the background noise less likely to be zero-mean and stationary than if the signal of interest is modeled as variable across trials.

It is therefore of great interest to be able to unveil a structure which is common among different realizations of a noisy signal, in spite of the variations in some aspects of this structure.

The analysis of single-trial data dates back to 1967, with Woody's cross-correlation averaging [3]. Several studies have proposed methods for such single-trial estimation, based on linear decomposition, Bayesian estimation or wavelet analysis [2], [1], [4], [5]. A direct denoising of EEG single-trial data with time-scale decomposition has been proposed, in which a wavelet template is designed, based on the average signal across trials [6]. The average signal is however potentially distorted by the inter-trial variability. In particular, this method is not appropriate for analyzing oscillations with inter-trial phase variability, as they tend to disappear in the average signal. Moreover, the chosen approach is not translation-invariant. More recently, a method has been proposed to automatically estimate the coefficients to be kept, using inter-trial statistics [7]. In another study, a translation-invariant wavelet transform is used in order to be more robust to temporal variability [4].

Time-frequency representations are commonly used to analyze the oscillatory nature of bioelectromagnetic signals. Because of above-mentioned inter-trial phase variability, oscillatory activity

which is not strictly phase-locked to the stimulus onset vanishes in the averaged data, while it is present in the averaged time-frequency power representation [8]. Such oscillatory activity, only visible in the averaged time-frequency power, is called *induced activity*, in contrast to *evoked activity* which is visible in the average signal.

It is interesting to use a data representation in which the structure can be described with few components ('sparse' representation). Doing so allows variability while controlling the robustness to noise. The Matching Pursuit algorithm allows the construction of such sparse representations. The principle of Matching Pursuit is to find iteratively within a redundant dictionary of "atoms" (i.e., elementary basis functions) the elements that can best describe a given signal [9]. The general principle is to iteratively subtract from the signal at each single trial its projection on the atom selected from the dictionary. Within a redundant dictionary, there can be more free parameters than with an orthogonal wavelet basis. In particular, with a Gabor dictionary, the amount of oscillation within an atom is free to change, allowing to describe both transient waves and sustained oscillations.

Analyzing biomedical signals with Matching Pursuit was pioneered by Durka and Blinowska on EEG transients [10]. Others applications of Matching Pursuit to biosignals include tracking epileptic seizures [11], analyzing breathing rates [12] or ECG data [13]. Matching Pursuit was adapted in [14] with a set of dictionaries whose structure was randomized to avoid statistical bias.

The present study is the first one to adapt Matching Pursuit to handle multi-trial data with cross-trial variability in all parameters (latency, frequency, width, phase). The originality of our method is to select the atom at a given iteration using an ensemble statistic that is robust to variability, and to subtract the atom in each trial by adapting its parameters: the subtraction step thus accounts for the inter-trial variability. Because the ensemble statistic is based on finding a consensus across trials, the method is called 'Consensus Matching Pursuit' (CMP).

The paper is organized as follows. Section II presents the problems that arise from using multi-trial data and the standard averaging procedure. Section III explains the general framework of Matching Pursuit, and section IV the multi-trial variants including CMP. Section V introduces the adaptation of CMP to the particular case of MEG and EEG. Section VI describes the synthetic and real datasets used for validation, and section VII gives the corresponding results. Finally, a discussion in section VIII highlights the main features of the method and gives ideas for further

developments.

II. Characteristics of multi-trial data

Classical Matching Pursuit (see III-C)aims at decomposing a single piece of data. In the present study, our goal is to analyze signals that can be segmented in sections, or 'trials', presenting similar events. Examples in the field of M/EEG are eye blinks, ECG waveform, epileptic spikes or evoked potentials (i.e., responses to a stimulation).

Typically, the signals display significant inter-trial variability, for example jitter in the latency of a given wave or of an oscillation, as illustrated in the toy example of Figure 1. The components of interest are moreover often buried in a very high noise, making a single occurrence of the signal insufficient for its interpretation.

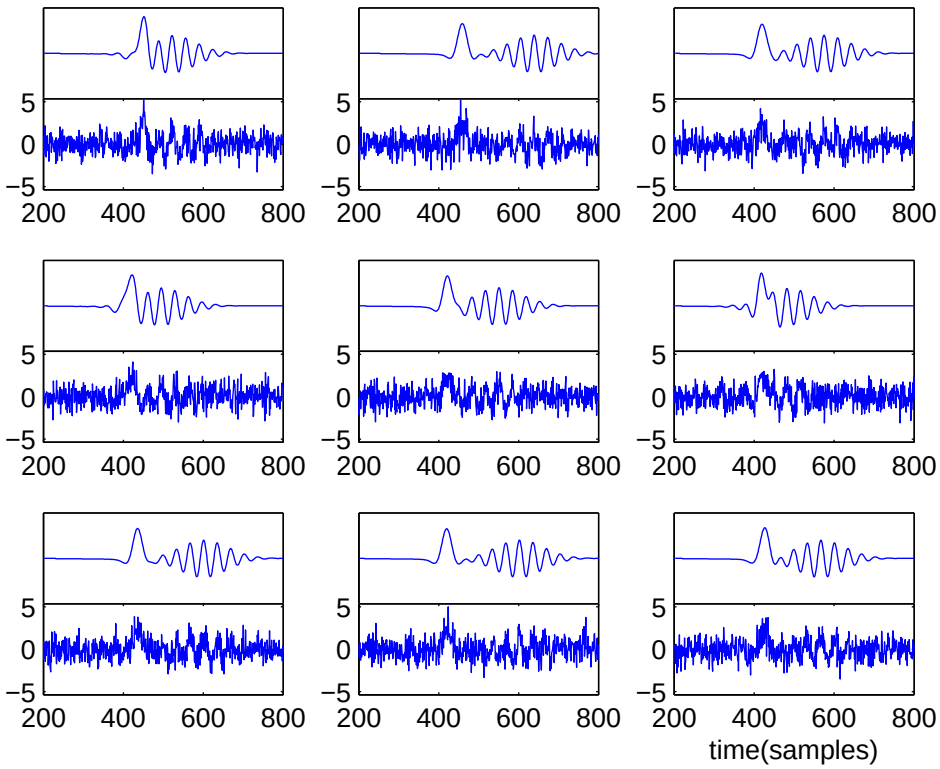


Fig. 1. A set of toy signals with inter-trial variability. The signal consists of a transient signal and an oscillation at 30 Hz. For each graph, the upper row is the noiseless data, and the lower row the noisy data on which the tests were conducted.

The poor signal-to-noise ratio (SNR) is the reason driving experimentalists to acquire a large number of trials, over which they perform statistics. Applying classical Matching Pursuit on each trial independently would be difficult, because such a procedure would hardly distinguish between the background noise and the signal of interest.

The common way to improve the SNR is to average the signals across trials, after they have been aligned to a common time-reference. Figure 2 shows the effect of averaging, with different choices of time-reference. In Figure 2(a), the signals are averaged across trials, with the time-reference of Figure 1, according to which there is variability in the latencies of both component. None of the components are correctly captured in the average: the oscillatory component almost disappears due to random phase variations across signals, and the shape of the wave-like component is distorted both in amplitude and in duration. In Figure 2(b), the signals are averaged after having been realigned on the latency of the earlier signal component: this transient component is well rendered but not the later, oscillatory, one. In Figure 2(c), the realignment is performed on the latency of the later signal component: the transient component almost disappears while the later component retains its true shape.

The Inter-Trial Coherency (ITC), or Phase Locking Factor (PLF) [15], is a measure of phase consistency across trials at a given frequency. As shown on the average in Figure 2 (middle), the ITC confirms that the oscillation in the original data is not phase-locked across trials.

The inter-trial variability is generally not consistent in time, making it undesirable to use a single time-reference over the signal duration. As shown in the above illustration, the time-reference should ideally be adapted to the signal components themselves - which are unknown a priori.

In order to avoid the phase cancellation effect of signal averaging (see (a) and (b) in Figure 2), Tallon-Baudry and Bertrand proposed in the mid-1990's to average the time-frequency power distribution of individual signals [15]. This representation allowed to observe "induced activity", i.e. time-frequency components which were not observable in the cross-trial average of the time-domain signals.

However, averaging time-frequency components with inter-trial latency variability leads to attenuation and spreading of the single-trial time-frequency representations, similarly to the attenuation and spread observed on the time-domain average. This effect can be seen when comparing the ERSPs of Figures 2(b) and 2(c).

Moreover, in a multi-channel framework, the time-frequency energy representation loses track of the polarities of dipolar fields; a representation with realigned components would be more easily interpretable in terms of underlying sources. A trial-to-trial tracking of the latency jitter of each component would permit such adaptive realignment.

The aim of Consensus Matching Pursuit (CMP) is to go further in analyzing multi-trial datasets, by employing statistics that are robust to trial-to-trial fluctuations and therefore permit to extract an undistorted representation of the single-trial events.

III. General Framework of Matching Pursuit

This article deals with multi-trial information extraction from biosignals, viewed as a set of time-dependent signals $s_k(t)$ indexed by trial number k. Trials can be determined either directly from the experimental protocol, or by segmenting the continuous data (visually or with a criterion based on correlation with a template). The data, measured simultaneously on several sensors, could be vector-valued, but for simplicity only scalar-valued signals are considered. The scalar signal $s_k(t)$ can for instance correspond (i) to a single sensor measurement (ii) to a single component after the measured data has been transformed by Principal Components (PCA) or Independent Components Analysis (ICA) or (iii) to the time-course of a source estimated by a source localization algorithm [16].

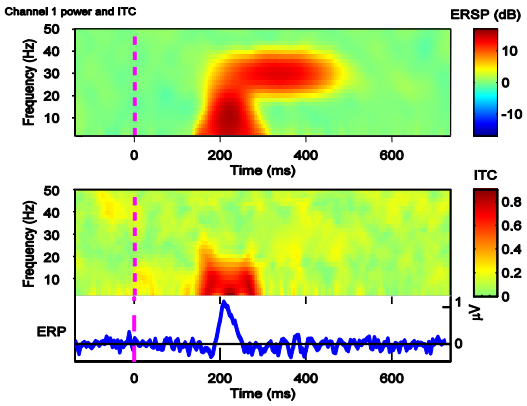
For the sake of simplicity, we describe our approach in the framework of continuous time, infinite duration, finite energy signals. Signals are thus modeled as square integrable (L^2) functions. The proposed Matching Pursuit methods rely on correlation, and the natural correlation measure between signals is the L^2 inner product

$$\langle f,\, g\rangle = \int f(t)\, g(t)^*\, dt\ , \quad f, g \in L^2 \tag{1}$$

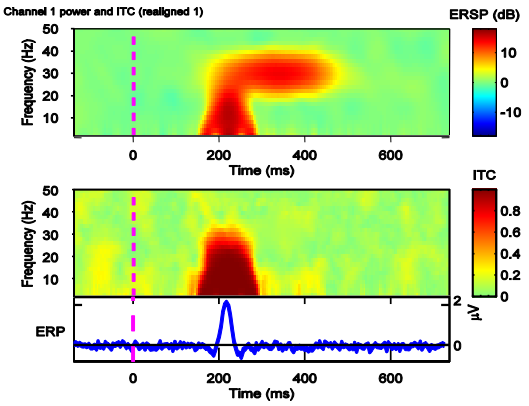
where * denotes the complex conjugate, and the corresponding norm is as usual denoted by $\|.\|$.

A. Signal modeling

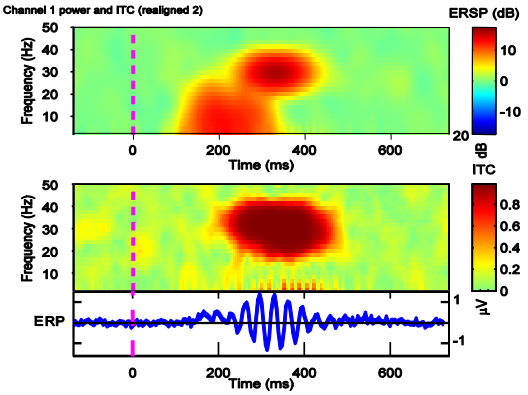
We consider that the relevant part of the signal consistently repeats over the trials, possibly with some variations in shape and latency. The remaining part, called “background activity”, is assumed to be stationary and uncorrelated between trials.



(a)



(b)



(c)

Fig. 2. Average signals and time-frequency maps of toy data (EEGLAB toolbox, with short term Fourier transform; window size: 125 ms). For each subfigure, the lower part is the average signal, the middle part the Inter Trial Coherency (i.e., the phase coherency across trials) and the upper part the Event Related Spectral Perturbation (i.e., power increase relative to the baseline). (a) Original signals. (b) Data realigned on the first component (transient wave) (c) Data realigned on the second component (oscillation). The trial-to-trial jitter produces a blurring of both the time-domain average signal (lower part of (a)) and the TF representation (upper part of (a)) compared to the data realigned on the transient (b) and on the oscillatory component (c). In the original time-domain average (bottom part of (a)), the oscillation almost completely disappears.

The signal $s_k(t)$ for a trial k is modeled as the superimposition of the background activity N_k and the relevant activity, constrained with a parametric model, as a sum of I parametric "atoms", whose parameters $(a_k, \mathbf{p}_k)$ are trial-dependent:

$$s_k(t) = \sum_{i=1}^{I} a_k(i)\psi_{\mathbf{p}_k(i)}(t) + N_k(t) \tag{2}$$

We make the hypothesis that variations in the parameters $\mathbf{p}_k(i)$ across trials are not too large, i.e. that they are of the order of magnitude of the size of the corresponding atoms.

The atoms $\psi_{\mathbf{p}_k}$ are elementary building blocks that are combined, with amplitude coefficients $a_k(i)$, to produce the relevant part of the signal. In the case of complex-valued atoms, the amplitude coefficients will also be complex-valued. Thus, for the typical case of real signals, only the real part of (2) will be of interest.

For a given trial k and atom index i, $\mathbf{p}_k(i)$ is a set of parameters defining the shape of the atom. The atoms are normalized so that $\|\psi_{\mathbf{p}}\| = 1$ for all $\mathbf{p}$ in the parameter space. The next subsection will focus on designing the set of all possible atoms in the representation, which is called the *dictionary*.

B. Design of dictionary

Our goal is to provide a robust way of estimating a set of atom parameters $\{\mathbf{p}_k(i)\}$ and their associated amplitudes $\{a_k(i)\}$. We will seek a sparse approximation of the signals within an *overcomplete dictionary*, i.e. a dictionary for which the decompositions are non-unique. Mathematically, this means that there exists a subset of parameters and amplitudes $\{a_k, \mathbf{p}_k | k \in K_0 \subset \mathbf{N}\}$ such that all a_k are not zero but

$$\sum_{k \in K_0} a_k \psi_{\mathbf{p}_k} = 0 \ .$$

Two important requirements for a dictionary are its *descriptive power*, i.e. its ability to represent the signals of interest with relatively few atoms, and its *interpretability*, i.e. that the parameters indexing the atoms convey information. Although overcomplete dictionaries do not provide uniqueness of decomposition, they have more descriptive power than more classical, orthogonal dictionaries[1]. Regarding interpretability, the choice of atoms and their parameters

[1]In an orthogonal dictionary, atoms are orthognal to each other: the inner product (1) of any pair of atoms is zero.

is motivated by the types of activities that will be encountered. The dictionary is furthermore supposed to depend continuously on the parameter space (even though this space will obviously be discretized). This constraint is imposed only for ease of presentation. However, the approach could easily be generalized, at the cost of some added complexity.

An example of such a dictionary is described in section V-A for the case of brain signals.

C. Classical Matching Pursuit

Given a dictionary of waveforms $\mathcal{D} = \{\psi_{\mathbf{p}}\}$, the corresponding *sparse regression* problem aims at finding signal expansions of the form $s(t) = \sum_{i=1}^{I} a_i \psi_{\mathbf{p}_i}(t) + N(t)$, keeping both the number I of atoms and the energy $\|N\|$ of the residual as small as possible. Such a problem is sometimes formulated as a variational problem, i.e. as the minimization of a penalty function of the form

$$\min_{I,\{a_i,\mathbf{p}_i,i=1,\dots I\}} \left(\int \left| s(t) - \sum_{i=1}^{I} a_i \psi_{\mathbf{p}_i}(t) \right|^2 dt + \lambda \sum_{i=1}^{I} |a_i| \right) ,$$

the last term favoring the sparsity of the expansion and λ being a tuning parameter. Despite recent progress, this so-called *Basis Pursuit Denoising* formulation (as well as several variants) generally leads to complex numerical optimization problems. Matching pursuit represents a valuable alternative, with lower complexity.

Matching Pursuit is an iterative method for decomposing a signal s in a dictionary. It seeks at each iteration the atom which best matches the signal and subtracts its contribution to the signal: at iteration 0, define $s^0 = s$ and for iteration $i \geq 0$,

$$s^{i+1} = s^i - \langle s^i, \psi_{\mathbf{p}(i)} \rangle \, \psi_{\mathbf{p}(i)}$$

and

$$\mathbf{p}(i) = \arg\max_{\mathbf{p}} |\langle s^i, \psi_{\mathbf{p}} \rangle|.$$

IV. Multi-trial Matching Pursuit

There are many ways to extend the scalar Matching Pursuit to a vector-valued Matching Pursuit that can handle multi-trial data. We present here three versions of such vector-valued MP: Multivariate MP (MMP), Induced activity MP (IMP) and Consensus MP (CMP).

A. Multivariate Matching Pursuit (MMP)

Multivariate Matching Pursuit, initially proposed in [17], and recently adapted to multi-trial, single-channel data [18], assumes that for a given iteration i the atom parameters are fixed across trials: $\mathbf{p}_k(i) = \mathbf{p}(i)$. Defining $s_k^0 = s_k$, for iteration $i \geq 0$, the residuals are computed according to

$$s_k^{i+1} = s_k^i - \langle s_k^i, \psi_{\mathbf{p}(i)} \rangle \, \psi_{\mathbf{p}(i)}. \tag{3}$$

In this way, the atom amplitude (defined by $\langle s_k^i, \psi_{\mathbf{p}(i)} \rangle$) is adapted to each trial.

The selection of the representative atom at iteration i can be done in a variety of ways. In the seminal paper [17],

$$\begin{aligned} \psi_{\mathbf{p}(i)} &= \arg\max_{\psi_{\mathbf{p}}} \left| \langle \textstyle\sum_{k=1}^K s_k^i, \psi_{\mathbf{p}} \rangle \right| \\ &= \arg\max_{\psi_{\mathbf{p}}} \left| \langle \overline{s}^i, \psi_{\mathbf{p}} \rangle \right| \end{aligned}$$

where $\overline{s}^i = \frac{1}{K} \sum_{k=1}^K s_k^i$ denotes the i-th residual of the signal, averaged over trials. Moreover it can be proved that the atoms of the MMP are the same as the atoms of a MP of the average signal $\overline{s} = \frac{1}{K} \sum_{k=1}^K s_k$.

The method performs well when the atoms $\psi_{p(i)}$ which correlate best with the average signal also correlate well with the individual signals. By considering the atoms of a matching pursuit decomposition of the average signal, and adapting the amplitude to the individual trials, the method accounts for amplitude variability across trials, but not for variability in the parameter space.

B. Induced activity Matching Pursuit (IMP)

Another way to select the representative atoms $\psi_{p(i)}$ consists in maximizing the average energy (or amplitude) across trials [19]:

$$\psi_{\mathbf{p}(i)} = \arg\max_{\psi_{\mathbf{p}}} \sum_{k=1}^K \left| \langle s_k^i, \psi_{\mathbf{p}} \rangle \right|$$

or

$$\psi_{\mathbf{p}(i)} = \arg\max_{\psi_{\mathbf{p}}} \sum_{k=1}^K \left| \langle s_k^i, \psi_{\mathbf{p}} \rangle \right|^2$$

As this is similar to the strategy for detecting induced activity in the time-frequency domain [8], we call this approach 'Induced activity Matching Pursuit' (IMP). Averaging the energy in the TF plane permits to detect activity that is not strictly phase locked from one stimulus to the other. A similar approach has been performed for multichannel data [20].

C. Consensus Matching Pursuit

In this paper, a less constrained version of multi-trial matching pursuit is proposed. It allows not only the atom amplitudes, but all the atom parameters to be *trial-dependent*, as in our model (2). We replace (3) with

$$s_k^{i+1} = s_k^i - a_k(i)\, \psi_{\mathbf{p}_k(i)} \ , \tag{4}$$

where the atoms $\psi_{\mathbf{p}_k(i)}$ depend on the trial k and the amplitudes $a_k(i)$ are given by

$$a_k(i) = \langle s_k^i, \psi_{\mathbf{p}_k(i)} \rangle \tag{5}$$

Our goal is to select a set of representative atoms $\psi_{\mathbf{p}_k(i)}$, by enforcing their *similarity across trials*. Each iteration consists in three steps:

- Select a large set of atoms independently at the level of each trial.
- Find a consensus atom that is most representative of atoms across trials.
- Subtract from each trial the atom that is most similar to the consensus atom, thereby tracking trial-to-trial variability.

We now detail the steps of the procedure, for an iteration i.

1) Selection of atoms for each trial: For each trial k, we compute the projection of the data $s_k^i(t)$ on all the atoms of the dictionary:

$$M_k^i(\mathbf{p}) = \left| \langle s_k^i, \psi_{\mathbf{p}} \rangle \right| \tag{6}$$

Atoms for each trial can be selected by a classical matching pursuit on the signal $s_k^i(t)$, as presented in III-C. In the present study we propose a simpler alternative. As the dictionary

depends on a continuous parameter space, we choose to construct a continuous map $M_k^i(\mathbf{p})$ and to extract all its local maxima:

$$L^i(k) = \left\{\mathbf{p} | M_k^i(\mathbf{p}) \text{ locally maximum at } \mathbf{p}\right\} \tag{7}$$

2) Find a consensus atom with a vote map: We hypothesize that there is a set of similar atoms that repeats consistently across trials. Our goal here is to find a global representative atom, or "consensus" atom, for such a set.

A simple choice would be to select the atom maximizing the induced activity, as presented in section IV-B. However, in case of jitter, this choice would lead to a global atom that is more spread out than the individual atoms across trials.

To retain the properties of the individual atoms, we propose to find the most prominent cluster of local maxima across trials. This could be performed by classical clustering algorithms. However, this would require determining the number of clusters, which is a difficult operation, not really needed here as we are only interested in one cluster at each iteration.

Instead, we use a weighted voting procedure, which is in similar to estimating a density of probability of atom parameters, from which one can find the most representative atom (i.e. the main mode of the density) and its spread in parameter space (i.e. the extent of the mode).

Each local maximum $\mathbf{p}_l$ votes in an accumulator map by adding to the accumulator a kernel C, centered on this local maximum. This results in a vote map

$$V^i(\mathbf{p}) = \sum_k \sum_{\mathbf{p}_l \in L^i(k)} M_k^i(\mathbf{p}_l)\ C(\mathbf{p}_l, \mathbf{p}) \tag{8}$$

The consensus atom $\overline{\mathbf{p}}(i)$ at iteration i is defined as that maximizing the vote map:

$$\overline{\mathbf{p}}(i) = \arg\max_{\mathbf{p}} V^i(\mathbf{p})$$

The spread of the vote for each local maximum is determined by a similarity measure $C(\mathbf{p}_l, \mathbf{p})$. Choosing for C a diagonal measure (for example $C(\mathbf{p}, \mathbf{p}') = \delta(\mathbf{p} - \mathbf{p}')$) would yield a vote in which the atom that has been selected most often is retained, and similar but slightly different atoms are not considered in the vote map. For the problem at hand, less localized choices should be favored, for instance the inner product modulus

$$C(\mathbf{p}_l, \mathbf{p}) = |\langle \psi_{\mathbf{p}_l}, \psi_{\mathbf{p}} \rangle|$$

or a Gaussian kernel centered at $\mathbf{p}_l$ in parameter space

$$C(\mathbf{p}_l, \mathbf{p}) = \exp\left(-(\mathbf{p}_l - \mathbf{p})^T \Sigma (\mathbf{p}_l - \mathbf{p})\right) \tag{9}$$

For computational reasons, we chose the latter similarity measure, using a block diagonal dispersion matrix Σ between parameters. More details are given after the description of the dictionary (section V-D).

3) Subtract atom for each trial: In this step, there is an interplay between the individual maps computed in step 1) and the global vote map of step 2). For each individual trial k, the atom in the individual map whose parameters are the most similar to $\overline{\mathbf{p}}(i)$ is selected. A natural similarity measure selects the atom that has most contributed to the vote map in (8):

$$\mathbf{p}_k(i) = \underset{\mathbf{p} \in L^i(k)}{\arg\max} M_k^i(\mathbf{p}_l)\ C(\mathbf{p}_l, \overline{\mathbf{p}}(i)) \tag{10}$$

However, such measure does not take into account the information on the spread of parameters that is contained in the vote map. Moreover, it requires to recompute the kernel C at each atom in $L^i(k)$. We therefore choose to use a similarity measure derived from the voting map:

$$\mathbf{p}_k(i) = \underset{\mathbf{p} \in L^i(k)}{\arg\max} M_k^i(\mathbf{p}) \hat{V}^i(\mathbf{p}) \tag{11}$$

with $\hat{V}^i$ mode of the vote map around the peak $\overline{\mathbf{p}}(i)$, as discussed in V-D.

The atom, denoted $\psi_{\mathbf{p}_k(i)}$, that combines a high amplitude and a high vote value is considered the most compatible with the voting map obtained on all trials. It is to be noted that this approach is conceptually similar to a maximum likelihood solution in a Bayesian framework.

The complex amplitude coefficient $a_k(i)$ corresponding to $\psi_{\mathbf{p}_k(i)}$ is estimated with (5). The contribution of the atom for each trial $a_k(i)\,\psi_{\mathbf{p}_k(i)}$ is then subtracted out from the signal $s_k^i(t)$, resulting in a new set of residuals $s_k^{i+1}(t)$, on which the next iteration proceeds.

Currently, the algorithm ends after extracting a predetermined number of atoms. A rough estimate of this number can be obtained from the number of blobs in a classical time-frequency analysis. Another option is to choose a large number of atoms, to compute a posteriori the contribution of the atoms to the energy of the original signal, and to discard atoms with a low contribution. Finding a better stopping criterion will be the focus of future work.

V. Application to brain signals

This section explains how the general framework presented in the previous section is implemented for the particular case of brain signals, such as electroencephalography (EEG) or magnetoencephalography (MEG). We provide our implementation of the method in MATLAB (Mathworks, Natick, MA) at http://www.insermU751.org/Software/CMP

A. Design of dictionary

In order to analyze both oscillatory activity such as gamma or alpha bursts, and lower frequency transient 'evoked potential' activity, an overcomplete wavelet representation of Gabor wavelets is chosen, defined in the time domain by:

$$\psi_\sigma^{f_0}(t) = (\pi\sigma^2)^{-1/4} e^{2i\pi f_0 t} e^{-\frac{t^2}{2\sigma^2}} \ .$$

and in the frequency domain by:

$$\Psi_\sigma^{f_0}(f) = (4\pi\sigma^2)^{1/4} e^{-\frac{\sigma^2}{2}(2\pi(f-f_0))^2} \ .$$

The scale σ stretches or compresses the time support of the wavelet, without modifying its frequency. The wavelet oscillates around the center frequency f_0, with a number of cycles of the order of the oscillation parameter (see Fig. 3):

$$\xi = 2\pi f_0 \sigma. \tag{12}$$

The optimal number of cycles depends on the type of activity under examination: generally lower for evoked, low-frequency activity and higher for oscillations.

An additional "latency" parameter u translates the atoms across time according to

$$\psi_\sigma^{f_0}(t-u). \tag{13}$$

Finally, an atom of the dictionary is represented by three parameters (u, f_0, σ), or (u, f_0, ξ). In this article, we choose the parameters $\mathbf{p} = (u, f_0, \xi)$ because the oscillation parameter ξ defined in (12) qualifies the transient vs. oscillatory nature of the activity.

In our overcomplete dictionary, the representation of a translated signal is the translation of the original signal representation. This translation-invariance property, which is highly desirable since signal components can have variable latencies across trials, would not be permitted by a

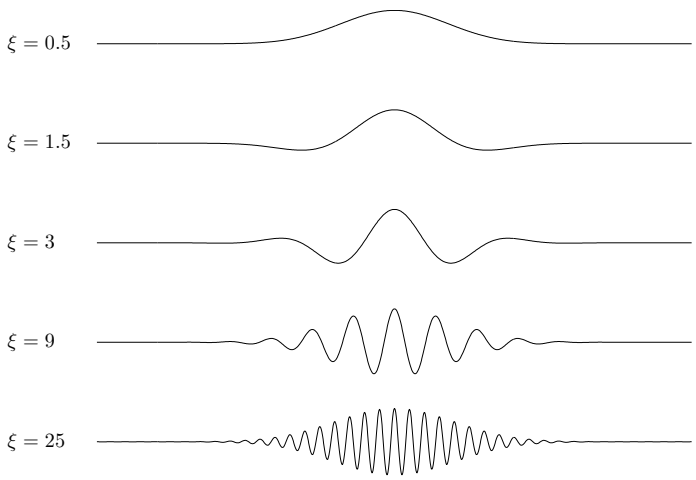


Fig. 3. Gabor atoms for different values of the oscillation parameter ξ (modified by varying f_0 with a contant σ, see 12). A low oscillation parameter produces a transient wave, and a high value a sustained oscillation.

wavelet orthogonal basis. The variability in the duration and frequency of activities is accounted for by the σ and f_0 parameters.

The representation with Gabor atoms is particularly interesting when there is a mixture of transient and oscillatory phenomena. Indeed, performing time-frequency analysis such as short-term Fourier transform or Morlet wavelets on a transient leads to a spread of energy towards high frequencies, which can intermingle with real oscillatory activity in the high frequencies. On the contrary, a more redundant representation can help in disentangling the two types of phenomena.

B. Time-frequency-ξ maps

As the dictionary presented in V-A is continuous with respect to the parameters, we choose to discretize the parameters and compute the projection of the signal on all the elements of the discretized dictionary. This results in time-frequency-ξ maps, which are extensions of the classical 2D time-frequency maps. The ranges of the frequency and ξ dimensions derive from a priori information on the range of interest. The discretization of ξ is heuristic, and in this paper, is set to $\{0.8; 1.5; 3; 5; 7; 9; 13; 25\}$. Further work will be necessary to define an optimal

sampling of this parameter.

For a given ξ and a particular window length, only a certain range of frequencies are acceptable, as the time-support of the atom (given by σ) must be smaller than the signal time window. Moreover, we choose not to analyze high frequencies ($f > 15$ Hz) for ξ's presenting few oscillations ($\xi < 2$): this stems from the assumption that high-frequency activity must be oscillatory. In the maps, all the frequencies bands that were not computed were set to zero.

C. Prewhitening

In our particular application of brain signals, the spectra of the signals have approximately a $1/f^{\alpha}$ shape, i.e. there is much more power in the low part of the spectrum. This has to be taken into account for the analysis of high frequency oscillations (above 15 Hz), as otherwise the low frequencies (evoked potentials, alpha rhythm) dominate the signal and its projections on wavelet or time-frequency atoms.

Several investigators have proposed to normalize the time-frequency representations by the average energy in the baseline, separately for each frequency [21], [15], [22]. We propose here a related method, spectral prewhitening, which consists in lifting the high frequencies in order to obtain an approximately flat spectrum. To do so, the single channel assumption that has been made up to now is momentarily relaxed and multi-channel recordings are considered.

The proposed method has two advantages. First, there is no more distinction between the baseline pre-stimulus activity and post-stimulus activity. This is interesting in the situations where the baseline can not be considered 'neutral' but rather comprises activity of interest [23]. Second, prewhitening can serve as a preprocessing step for multi-channel blind source separation, such as ICA, which would help identifying high-frequency sources, otherwise masked by low-frequency activity [24].

The prewhitening procedure is the following. First, a fast Fourier transform (FFT) is performed at each trial and each channel of the recording, with a Tukey window. Second, the spectrum is computed by averaging over channels and trials the squared amplitude of the FFTs. As this is performed over all channels, it is assumed that the phenomena of interest will not be dominant in the spectrum, and that the procedure will instead give an more general estimate of the general frequency shape. Third, a prewhitening function $P(f)$ is computed as the inverse of the spectrum on a window of interest, taken here as 1-100 Hz. This window is used in order to avoid lifting

up high-frequency noise, in frequency bands of no interest or at frequencies above the cut-off frequency of an anti-aliasing filter.

Each FFT of each trial in each channel is then multiplied by the prewhitening function $P(f)$, and the inverse FFT permits to come back to the time-domain.

D. Details of voting procedure

As mentioned in (9), the similarity function in the vote map (8) is a Gaussian kernel in the parameter space. Consequently, the accumulator is a piece of a 3D $time \times frequency \times \xi$ space (with the same discretization and parameter ranges as for the single-trial maps). Each peak of each trial 'votes' in the accumulator. The vote is performed by adding to the accumulator a Gaussian kernel centered on the peak and weighted by the peak's value.

The Gaussian kernel is:

$$C(t,f,\xi) = \exp\left(-\frac{1}{2}\left(\frac{t-t_{peak}}{S_t(t,f,\xi)}\right)^2 + \left(\frac{f-f_{peak}}{S_f(t,f,\xi)}\right)^2 + \left(\frac{\xi-\xi_{peak}}{S_\xi(t,f,\xi)}\right)^2\right), \qquad (14)$$

with $(t_{peak}, f_{peak}, \xi_{peak})$ coordinates of the peak in the 3D map and (S_t, S_f, S_ξ) extension parameters of the kernel in the three dimensions. The extension parameters are chosen as twice the full-width half maximum (FWHM) of the time-frequency atom $\psi(t,f,\xi)$ along the time and frequency dimension, and one bin in the ξ dimension.

This voting procedure is similar to the construction of the time-frequency maps from MP atoms proposed by Durka and colleagues [25], with the difference that we keep the 3 dimensions (t,f,ξ) instead of collapsing along the ξ dimension.

The region around the maximum of the voting map is extracted by fitting to it a 3D Gaussian kernel. This results in a new voting map $\hat{V}(t,f,\xi)$ with a single mode (used in (10)). This permits to avoid the influence of secondary modes of the voting map corresponding to other components of the signal.

VI. Validation method

Our goal is to test the recovery of the time-frequency structure of the data by the CMP algorithm despite variability in latency across trials, for both transient evoked potentials/fields and high-frequency oscillatory activity. The CMP is compared with two other variants of multivariate MP, the MMP, and the Induced MP (IMP), which were introduced in section IV.

In a first step, the algorithms are applied to the toy data shown in Figure 1, with added white Gaussian noise (SNR = 0 dB).

In a second step, the algorithms are applied to two sets of 50 trials of realistic synthetic data with a mixture of evoked potentials and 40 Hz oscillations at a realistic SNR. The difference between the two synthetic datasets is that only the second one has time jitter across trials.

Finally, the methods are tested on two real datasets: intracerebral recordings for a visual recognition protocol, and EEG recordings of epileptic discharges.

The details of the datasets are presented below. Dipole simulation was performed with the Fieldtrip toolbox[2]; the EEGLAB toolbox [26] was used for data handling. Prewhitening as described in V-C was performed for the realistic simulations and the visual task data, which presented low-amplitude high-frequency oscillations.

A. Toy data

50 trials of the toy data described in Figure 1 have been generated. The parameters of the simulated atoms are: (t,f,ξ)= (200 ms, 10 Hz, 1) and (250 ms, 30 Hz, 11). The amplitude, the ξ parameter and the frequencies of the atoms remain constant across trials. A latency jitter is applied to the two atoms with a standard deviations of three times the atom width. White Gaussian noise was added to the data, with a standard deviation set to the maximum amplitude of the simulated (noiseless) data.

B. Realistic synthetic data

Two sets of 50 trials (SimA and SimB) have been generated. Each of these datasets comprises four spatio-temporal components, each component being the product of a topography and a time-course. The topographies were generated using electrical dipoles simulated within a sphere, with a montage comprising 83 electrodes (10-10 system). The relevant topographies for the present study are (i) two dipoles in the posterior (occipital) region (maximum of the fields at electrodes O1/PO3 and O2/PO4 respectively), (ii) a dipole in the parietal region (maximum of the field at electrode Pz), (iii) a low occipital topography generated by one dipole.

[2]available at urlhttp://www.ru.nl/fcdonders/fieldtrip

The time-course of the first component comprises an evoked potential with two waves, peaking at 100 ms (P1) and 170 ms (N170). The time-course of the second component is a single wave peaking at 300 ms in average (P300). The third is a 35 Hz oscillation (late gamma) peaking at 300 ms in average: (t,f,ξ)=(300,35,5). Fluctuations have been introduced in amplitude for all components, and in width for the P300 and the late gamma activity.

In SimA, there is no latency jitter; in SimB, a jitter in latency is introduced for both the P300 wave (270 to 500 ms with a skewed distribution) and the late gamma oscillation (250-350 ms). The P300 and gamma jitters are independent of one another.

Realistic noise is obtained by placing dipoles randomly within the sphere, with a random time course (Gaussian white noise). This results in noise with realistic spatial correlation. Temporal correlation is obtained by distorting the spectrum of the noise at each electrode. The realistic spectrum and the average energy across channels (measured as the mean variance) are obtained on one sample of real data. In addition, alpha waves are simulated by assigning to two occipital dipoles sine waves oscillating at 10 Hz, with an envelope function. The envelope is constant over the baseline, and returns to zero between 100 and 200 ms after the trigger time. The phases are random across trials, and uncorrelated between the two dipoles, resulting in a fluctuation of amplitude of the simulated alpha across trials.

The waveforms of the simulated datasets are presented in Figure 4, before prewhitening.

C. Real data

We used two datasets for the validation on real data. The first dataset consisted in intracerebral EEG data (stereotaxic EEG, SEEG) obtained in a visual task. The robustness of the method to latency jitter was tested by artificially introducing a jitter in the data across trials. The second dataset consisted in epileptic discharges measured on scalp EEG.

1) Visual task: This data consists in intracerebral EEG recordings in an epileptic patient, sampled at 1kHz. The intracerebral electrodes had been implanted in the occipital region for the sole clinical purpose of presurgical evaluation. The patient performed a variation of the protocol introduced by Tallon-Baudry and colleagues [15], which consists in observing a succession of illusory (curved and non curved), real triangles and no-triangle stimuli. The variation consisted in presenting the stimuli triangles pointing up or down. The patient was instructed to respond to the curved illusory triangles with a button press. Such a protocol has been shown to result

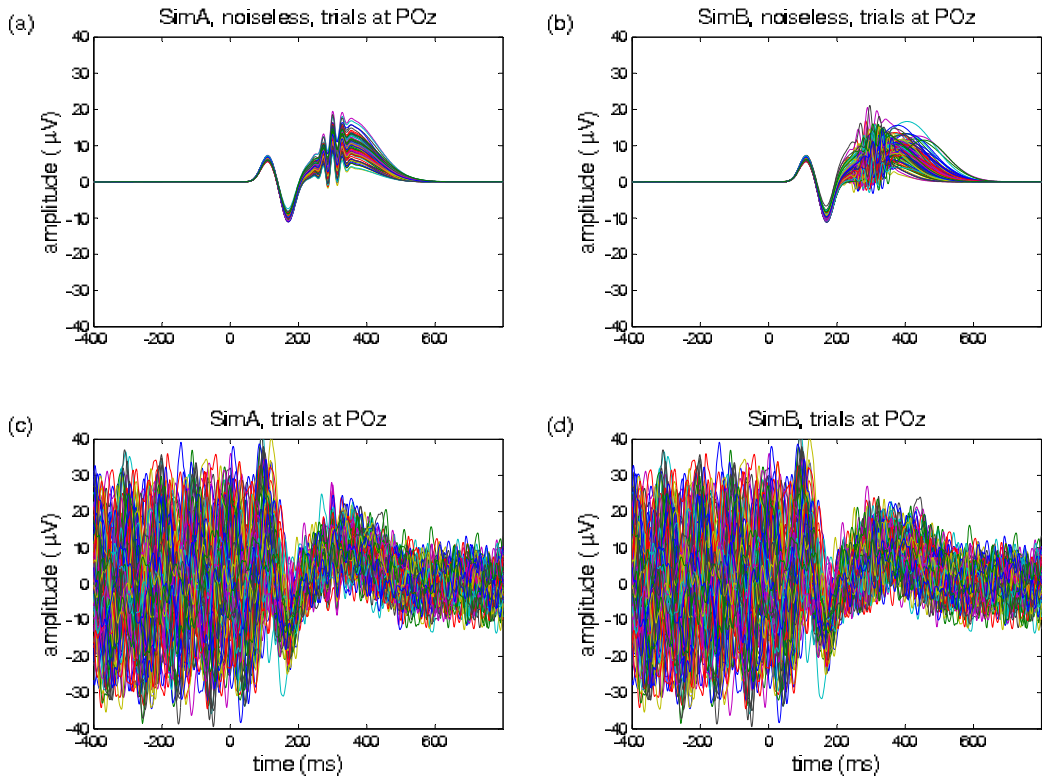


Fig. 4. Simulated waveforms across trials, at electrode POz. (a) SimA (no temporal jitter) without noise. (b) SimB (with temporal jitter) without noise. The jitter was tuned independently for each component of the signal. (c) and (d) Same signal with added noise (stationary background plus alpha waves) .

in a mixture of event-related potentials and high-frequency oscillatory activity [15]. Here, only two of the stimuli are analyzed: illusory triangles and real triangles, both pointing upward.

The chosen channel is located in the occipital cortex, and was selected because this contact exhibited high frequency activity in a time-frequency analysis. In a first step, the method was applied to the original data. In a second step, the data was jittered in order to test the robustness of the reconstruction: at each trial, a random delay has been introduced on the whole time window, with a Gaussian distribution, a mean of 100 ms and a standard deviation of 33 ms.

The time-frequency analysis (event-related spectral perturbation, ERSP) [21] and the phase-locking factor (or inter-trial coherency, ITC) [15] are shown in Figure 5. They were computed on the original signal (before prewhitening), with the EEGLAB toolbox [26]. The ERSP was performed with a short-term Fourier transform, with a window size of 256 ms (256 points), with a normalization with respect to the baseline (log-transform of the ratio between the squared amplitude at each point of the time-frequency plane and the mean energy of the baseline). The

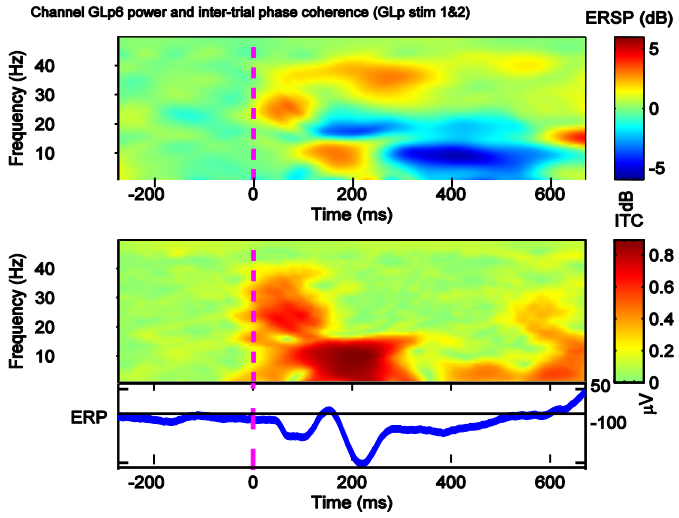


Fig. 5. Classical time-frequency analysis of the real data (without jitter). Upper panel: energy relative to the baseline (or event-related spectral perturbation, ERSP), computed with a short-term Fourier transform. Lower panel: phase locking factor (or inter-trial coherency, ITC). Three main blobs with power increase are seen: (i) around (60 ms,25 Hz), which corresponds to evoked activity as shown in the ITC plot, (ii) around (250 ms, 35 Hz), which corresponds to induced gamma, and (iii) at (170 ms, 10 Hz), reflecting the evoked potential.

ITC is a measure of the phase consistency across trials at a given point of the time-frequency plane, which permits to distinguish between evoked and induced activity [8].

2) Epileptic oscillations: This data comes from an overnight recording on a epileptic patient. The recording presented several epileptic discharges consisting in a high frequency oscillation (around 30 Hz) followed by a biphasic wave. We visually selected 23 events; sections of 2 seconds around the event were created.

Three trials of the real data are presented in Figure 6, along with their time-frequency representations, which shows oscillations around 27 Hz.

VII. Results

A. Toy example

The time-frequency-ξ map (used in the Induced MP) and the vote map (used in the Consensus MP) are shown in Figure 7, for iteration 1 (see section (VI-A) for details on the simulation). Because of the temporal jitter, for IMP, there is an overestimation of the width of the atoms

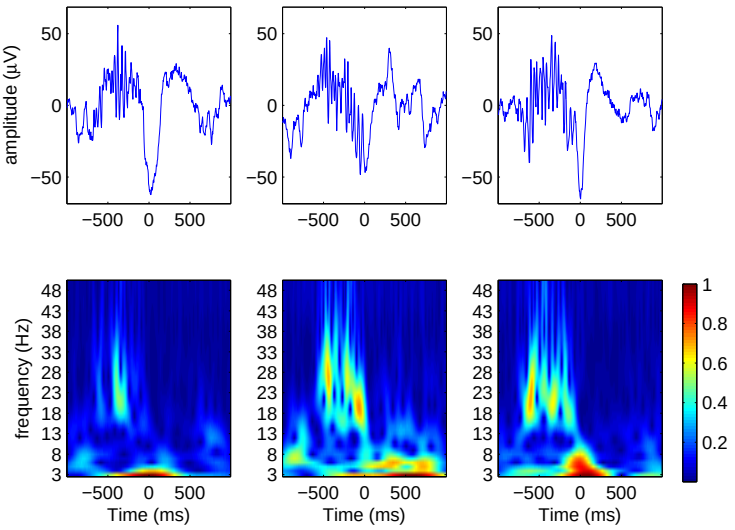


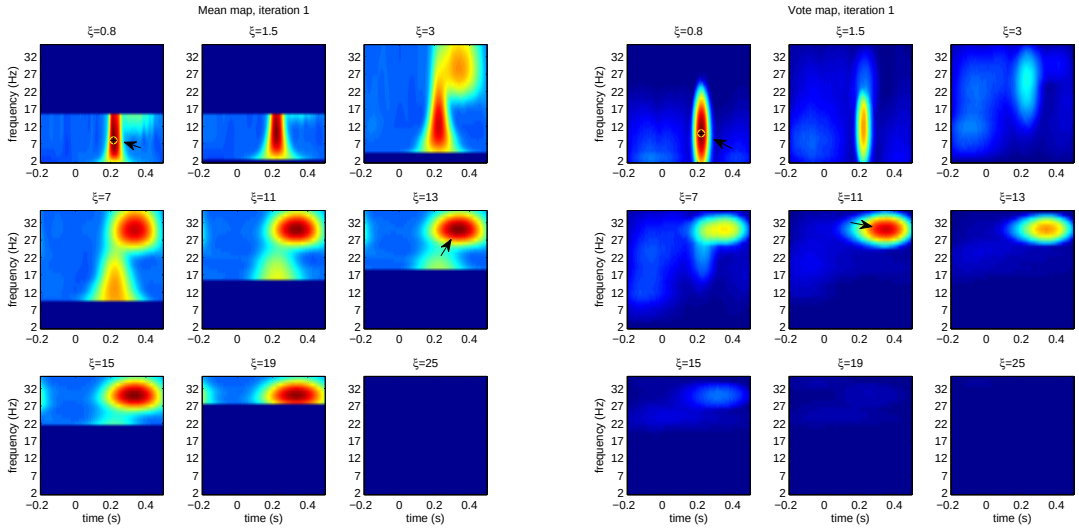
Fig. 6. Epileptic discharges. Top graph: scalp EEG signals (three events out of 23). Each discharge is composed of an epileptic oscillation and a biphasic wave. The channel used for the analysis is C3-P3. Bottom graph: time-frequency transform ($\xi = 5$). The oscillation is visible at $f = 27$ Hz

when performing the mean, as had been seen in Figure 2). This is reflected in a shift upward along the dimension ξ: indeed,for a fixed frequency, ξ is proportional to the width as shown in (12).

Figure 8 presents for CMP the local peaks found in the time-frequency-ξ maps for all the trials, for iteration 1. The identified peaks form a cluster around the peak detected in the vote map. There are many spurious peaks due to noise, but one can note that the presence of signal clears the region around the consensus peak.

Figure 9 shows the reconstructed signal obtained by combining the first two single-trial atoms detected by the CMP. The noiseless signal was reconstructed with a very good accuracy.

In Figure 10, the atoms detected by the three methods introduced in section IV are compared: MMP (on average data), IMP (on average time-frequency) and CMP. For each method, the atoms are scaled by the median fitted amplitude across trials. The length of the transient wave in MMP and of the oscillation in IMP are overestimated because of the averaging procedure (respectively in time domain and frequency domain). The CMP was able to recover the shape of the original signals, which is visible on the mean on realigned data.



(a) (b)

Fig. 7. Time-frequency-ξ maps for the toy simulation, for iteration 1 out of 2: mean induced map in 7(a) and vote map in 7(b). Frequency bands that were not computed are set to zero (see section V-B). The two local maxima are indicated by an arrow: there is one peak in the low ξ that corresponds to the transient wave, and another for a high value, reflecting the oscillation. The oscillation peak is shifted upwards to $\xi = 13$ on the mean map , whereas on the vote map it remains at the correct location ($\xi = 11$).

B. Realistic synthetic data

Figure 11 presents the results of the three MP algorithms on the two realistic simulations: without jitter across trials (SimA) and with jitter (simB). The atoms are represented at their positions in the time-frequency plane and the oscillation parameter ξ is visible from their time courses. All methods were able to recover the evoked potential (low frequency atoms); the MMP explained the wave with two consecutive atoms, which is a less sparse but more flexible representation, whereas the IMP and CMP used one atom only, with higher oscillation parameter ξ. Indeed, a wave with a ξ of 3 is a good approximation of two consecutive waves.

Only the IMP and CMP were able to recover the alpha wave in the background, as it is not phase-locked across trials and therefore not visible on the temporal mean. The MMP recovered the high frequency oscillation on SimA (where it is phase-locked) but not on SimB. IMP and CMP found one atom for the high frequency oscillation on both SimA and SimB; the oscillation

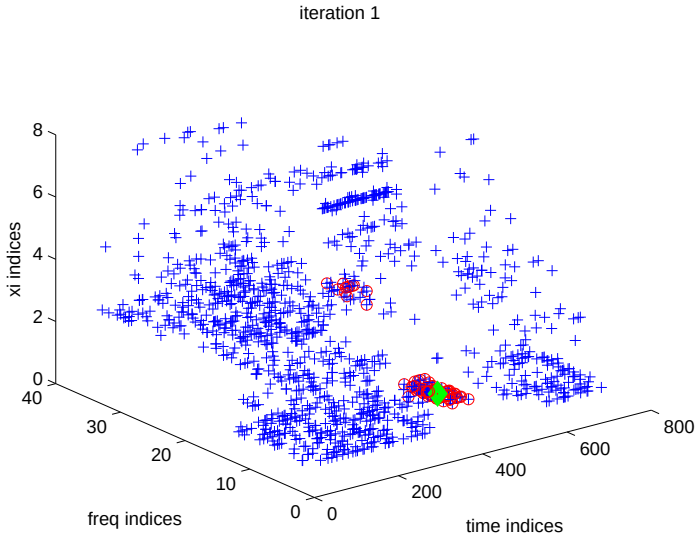


Fig. 8. Local peaks across all trials at iteration 1 of CMP on the toy data (atom corresponding to the oscillation). The consensus peak of the vote map is shown (green diamond). The peaks that are most similar to this consensus are highlighted (red circles). Most detected peaks (in red) form a cluster around the peak detected in the vote map.

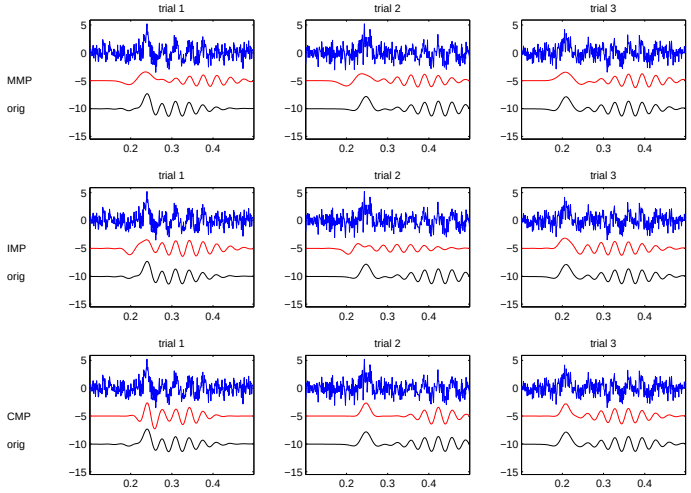


Fig. 9. Toy data: signal reconstructed by combining the first 2 atoms detected with the CMP method, for the first 9 trials. For each trial, the noisy signal is presented in the top row, the reconstructed signal in the middle row and the noiseless signal in the bottom row. The CMP was able to recover the noiseless signals with a high accuracy.

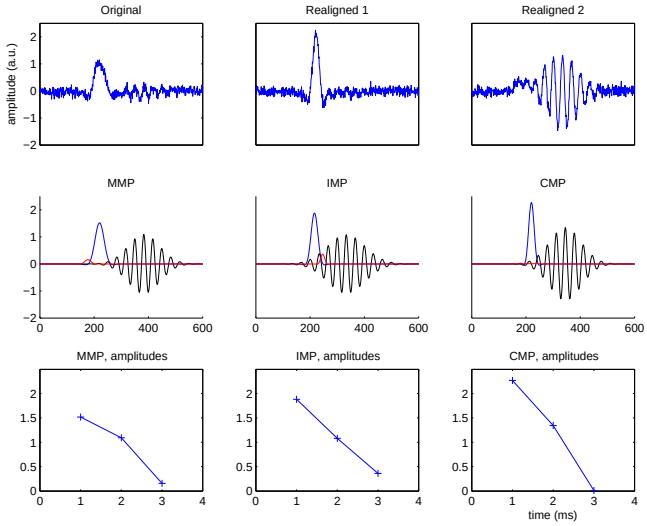


Fig. 10. Comparison of the results of the three algorithms on the toy data.**Top row:** average data across trials. (left) Original data, (middle) realigned on the transient wave (right) realigned on the oscillation.**Middle row:** detected atoms. (left) MMP, on average data, (middle) IMP, on average time-frequency representation, and (right) CMP. The (t, f, ξ) parameters are obtained from the summary statistic (mean or vote); the amplitude is the median amplitude across trials.**Lower row:** median amplitude of the three atoms. The CMP was able to recover the shape of the original signals, as visible on the mean on realigned data. The CMP is more sparse: the amplitude of the third atom is close to zero, whereas MMP and IMP required a third atom to represent the dataset.

parameter was slightly overestimated for SimB in the IMP, whereas CMP found the exact same representation for the phase-locked (SimA) and jittered (SimB) data.

In Figure 12, examples recovered parameters across trials are shown for CMP. There is a strong correlation between the simulated and recovered parameters for latency, frequency, ξ of P300 and gamma oscillation, and for the amplitude of the P300. The correlation is lower for amplitude of the gamma oscillation, whose estimation was less robust, due to the high noise level.

C. Real data

1) Visual task: Figure 13 compares the atoms obtained on the real data to the atoms obtained on the jittered data (as described in the section VI-C), for the three variants of MP.

For the original data, all methods identify similar atoms for describing the evoked potential (frequencies < 10 Hz). Both IMP and CMP identify an oscillation around 30 Hz, which was

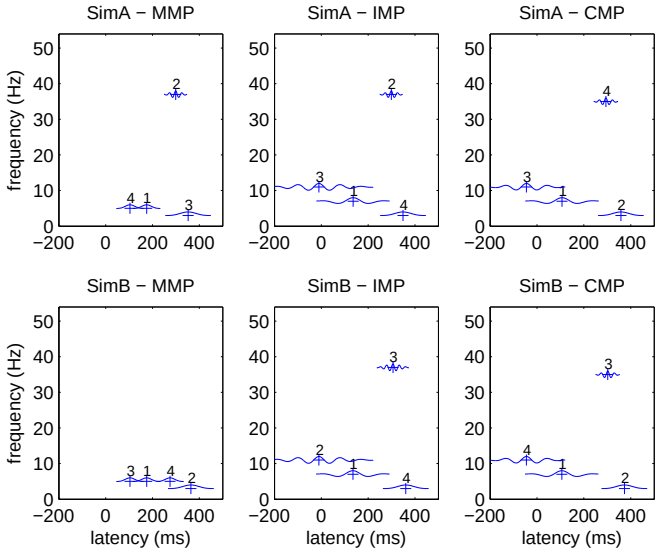
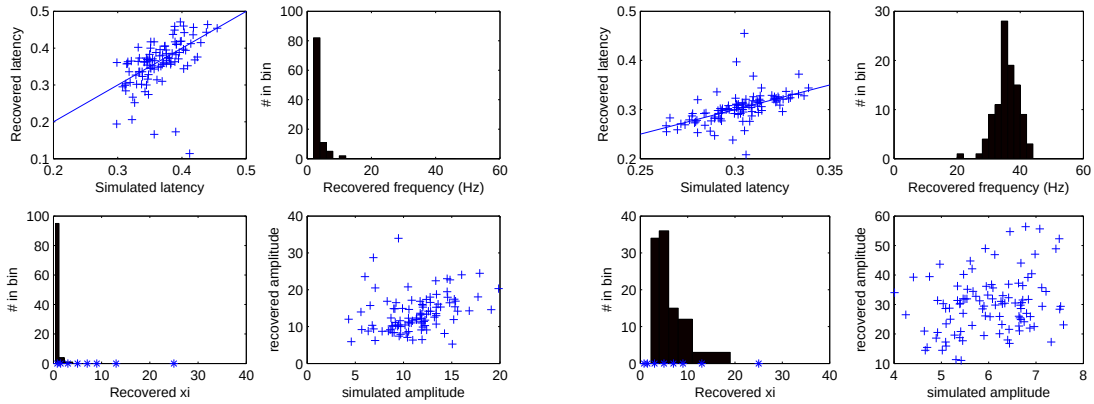


Fig. 11. Comparison of the results of the different matching pursuit procedures.**Top row**: SimA (no jitter);**Bottom row**: SimB (with jitter). Without jitter, all methods are able to recover the oscillation (atom 2 for MMP and IMP, atom 4 for CMP). With jitter, the oscillation disappears from the average and is not visible by MMP.



(a) P300 (b) late gamma oscillation

Fig. 12. Realistic simulation with jitter: parameters recovered by the CMP, for the atoms corresponding to the P300 and the gamma oscillation. There is a strong correlation between the original and the recovered parameters. Note that the amplitude of the recovered atom is not to scale because of the prewhitening procedure.

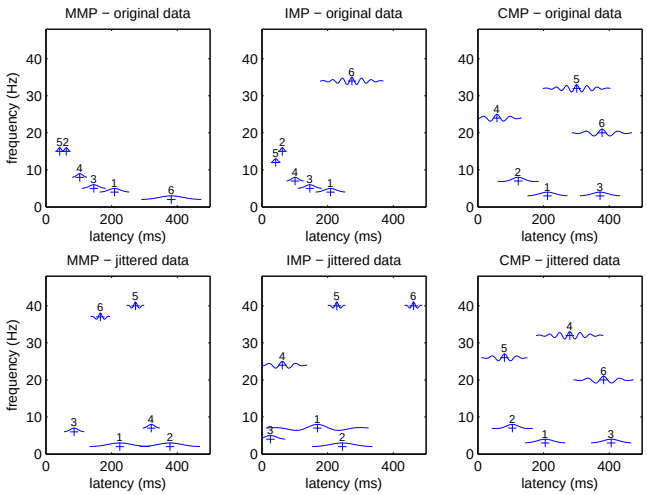


Fig. 13. Comparison of the atoms found on the real data (upper panel) and the atoms found on the data with jitter (lower panel). For the CMP, the representations are very similar, showing that the decomposition was robust to the added jitter.

not visible on the temporal mean ('induced gamma'). The CMP has detected one oscillation at 400 ms and 20 Hz. Note also that IMP has missed this 20 Hz component because of its less sparse representation.

For the jittered data, only the CMP is able to recover an identical decomposition; both MMP and IMP have been highly distorted.

2) Epileptic discharges: For the epileptic discharges, prewhitening was not applied because the oscillation has large amplitude, and also because we wanted to investigate the reconstruction of the signal based on the detected atoms. We found that a large number of atoms was necessary to reconstruct the signal faithfully (as seen in what follows), and performed the decomposition with 20 atoms.

The detected atoms are shown in Figure 14, for the three MP methods. All methods identify a cluster of atoms in the range -800 to 0 ms, between 15 and 30 Hz. The reconstructed signal with the 20 CMP atoms are presented in Figure 15.

The goodness of fit of the reconstructed data is presented in figure 16, for different number of atoms used in the reconstruction. The curve for CMP is consistently above that of the IMP, confirming that the CMP requires less atoms (i.e. is more sparse) for describing a signal with fluctuations.

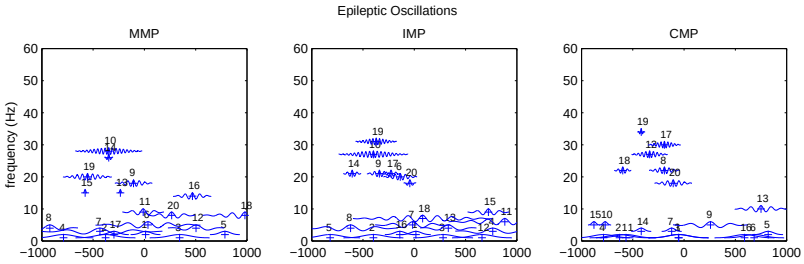


Fig. 14. Atoms found on the epilepsy EEG data, with the three MP methods. All methods identify oscillations in the range -800 to 0 ms, between 15 and 30 Hz.

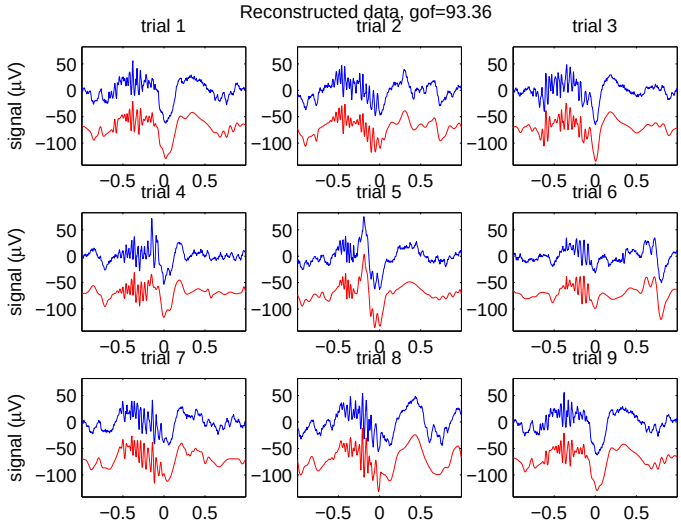


Fig. 15. Reconstructed data from CMP parameters (bottom curve of each trial) along with the original data (top curve), for nine trials out of 23.

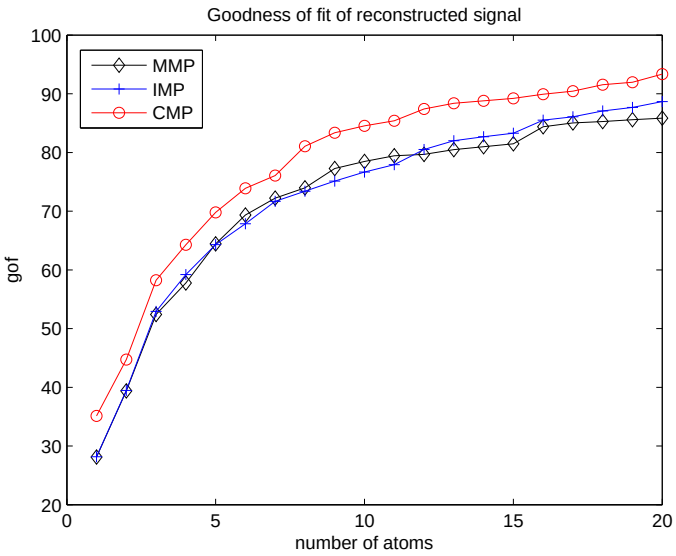


Fig. 16. Goodness of fit (GOF) of the reconstructed data, for a variable number of atoms. The CMP is consistently above IMP and MMP, meaning that its decomposition is more sparse for a given GOF.

VIII. Discussion

The method proposed here, Consensus Matching Pursuit (CMP), allows the determination of the atoms that can sparsely describe a multi-trial M/EEG dataset. Such atomic decomposition can handle both evoked potentials and high-frequency activity, and the resulting atoms are useful for the detailed characterization of a dataset. In particular, it permits to (i) determine whether a blob in the classical time-frequency representation is due to an actual oscillatory phenomenon or to a transient and (ii) estimate the actual width of a phenomenon (wave or oscillation) despite the inter-trial variability.

In the presence of fluctuations (i.e., jitter) across trials, the CMP gives a more faithful representation of the underlying atoms than a simple mean in time or in time-frequency, as used in the MMP and IMP. Indeed, the mean is prone to distortion: for example, in the case of a fixed pattern with jitter across trials, as presented in our toy example, the mean across trials is in fact the convolution of the true pattern by the distribution of latencies. On the contrary, Consensus Matching Pursuit aims at detecting the actual single-trial patterns, and thereby unveil the structure blurred by the mean. Moreover, subtracting atoms with trial-dependent parameters permits to better suppress the corresponding fluctuating activity at a given iteration, resulting in

a more sparse representation.

The method can serve as a preprocessing step that enables to find basis functions, which can then be used in order to parameterize the dataset for a more advanced dedicated single-trial analysis [27]. Other parameterized dictionaries could be used, for example chirplets that are well-fitted to the shape of the evoked potentials [28]. It is also possible to use a mixture of dictionaries, or a more general dictionary without parameterization. In the latter case, the atoms for each trial could not be selected as local peaks in a map, but rather with a classical matching pursuit.

Future steps will include a more comprehensive study of (i) the optimal sampling of the parameter space (ii) the statistical properties of the parameters: statistical significance of a given atom, covariance of the parameters, trial peaks clustering and (iii) the multi-channel nature of the dataset [5].

For the statistical analysis, we could rely on bootstrap techniques [29], [30]. For the multi-channel aspect, Consensus Matching Pursuit could be performed at each sensor, which would all vote for the best atom at a given iteration [19]. There should be less temporal variability for multi-channel than for multi-trial, even though phase differences across sensors could appear because of coupled oscillators or traveling waves.

A spatial constraint could be useful in multi-channel processing [20], [31]. Another interesting option would be to operate at the level of reconstructed sources. Source localization could help improve the sparsity of the representation, and directly give information of the observed phenomena from the sources within the cortex.

ACKNOWLEDGMENT

The authors would like to thank Catherine Liégeois-Chauvel and Agnès Trébuchon-Da Fonseca for the real data, Jean-Marc Lina and Boris Burle for useful discussions on the method, and the anonymous reviewers for useful suggestions.

REFERENCES

[1] A. Holm, P. Ranta-aho, M. Sallinen, P. Karjalainen, and K. Muller, “Relationship of P300 single-trial responses with reaction time and preceding stimulus sequence,” vol. 61, no. 2, pp. 244–252, 2006.

[2] T. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. Sejnowski, “Analysis and visualization of single-trial event-related potentials,” vol. 14, pp. 166–185, 2001.

[3] C. Woody, “Characterization of an adaptive filter for the analysis of variable latency neuroelectrical signals,” Medical and Biological Engineering, vol. 5, pp. 539–553, 1967.

[4] Z. Wang, A. Maier, D. Leopold, N. Logothetis, and H. Liang, “Single-trial evoked potential estimation using wavelets,” vol. 37, no. 4, pp. 463–473, Apr 2007.

[5] P. Ranta-aho, A. Koistinen, J. Ollikainen, J. Kaipio, J. Partanen, and P. Karjalainen, “Single-trial estimation of multichannel evoked-potential measurements,” vol. 50, no. 2, pp. 189–196, 2003.

[6] R. Quian Quiroga and H. Garcia, “Single-trial event-related potentials with wavelet denoising,” vol. 114, no. 2, pp. 376–390, 2003.

[7] M. Benkherrat, B. Burle, S. Allain, T. Hasbroucq, and F. Vidal, “Individual evoked potential extraction by multiresolution wavelets decomposition,” in Proceedings EUROCON2005, 2005.

[8] C. Tallon-Baudry and O. Bertrand, “Oscillatory gamma activity in humans and its role in object representation.” vol. 3, no. 4, pp. 151–162, 1999.

[9] S. Mallat and Z. Zhang, “Matching pursuit with time-frequency dictionaries,” vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[10] P. Durka and K. Blinowska, “Analysis of EEG transients by means of matching pursuit,” Ann. Biomed. Eng., vol. 23, pp. 608–611, 1995.

[11] P. Franaszczuk, G. Bergey, P. Durka, and H. Eisenberg, “Time-frequency analysis using the Matching Pursuit algorithm applied to seizures originating from the mesial temporal lobe,” vol. 106, pp. 513–521, 1998.

[12] M. Akay and E. Mulder, “Investigating the effect of maternal alcohol intake on human fetal breathing rate using adaptive time-frequency analysis methods,” vol. 46, no. (1-2), pp. 153–164, Sep 1996.

[13] H. Sava, P. Pibarot, and L. Durand, “Application of the matching pursuit method for structural decomposition and averaging of phonocardiographic signals,” vol. 36, no. 3, pp. 302–308, May 1998.

[14] P. Durka, D. Ircha, and K. J. Blinowska, “Stochastic time-frequency dictionaries for matching pursuit,” IEEE Transactions on Signal Processing, vol. 49, no. 3, pp. 507–510, 2001.

[15] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, “Stimulus specificity of phase-locked and non-phase-locked 40 hz visual responses in human.” vol. 16, no. 13, pp. 4240–4249, July 1996.

[16] S. Baillet, J. Riera, G. Marin, J. Mangin, J. Aubert, and L. Garnero, “Evaluation of inverse methods and head models for EEG source localization using a human skull phantom,” vol. 46, pp. 77–96, 2001.

[17] P. Durka, A. Matysiak, E. Martinez-Montes, P. V. Sosa, and K. Blinowska., “Multichannel matching pursuit and eeg inverse solutions.” Journal of Neuroscience Methods., vol. 148, pp. 49–59, 2005.

[18] R. Sieluzycki, C.and Konig, A. Matysiak, R. Kus, D. Ircha, and P. Durka, “Single-trial evoked brain responses modelled by multivariate matching pursuit,” in press, to appear.

[19] H. Gribonval, R.and Rauhut, K. Schnass, and P. Vandergheynst, “Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms,” IRISA, Tech. Rep. 1848, May 2007.

[20] D. Studer, U. Hoffmann, and T. Koenig, “From EEG Dependency Multichannel Matching Pursuit to Sparse Topographic Decomposition,” Journal of Neuroscience Methods, vol. 153, no. 2, pp. 261–275, 2006.

[21] S. Makeig, “Auditory event-related dynamics of the eeg spectrum and effects of exposure to tones.” vol. 86, no. 4, pp. 283–293, Apr. 1993.

[22] J. Lachaux, E. Rodriguez, J. Martinerie, C. Adam, D. Hasboun, and F. Varela, “A quantitative study of gamma-band activity in human intracranial recordings triggered by visual stimuli.” vol. 12, no. 7, pp. 2608–2622, July 2000.

[23] C. Tallon-Baudry, M. Henaff, J. Isnard, and C. Fischer, “Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus.” vol. 15, no. 5, pp. 654–62, May 2005.

[24] C. Bénar, T. Papadopoulo, and M. Clerc, “Topography-Time-Frequency atomic decomposition for event-related M/EEG signals,” pp. 5461–5464, Aug. 2007.

[25] P. Durka, J. Zygierewicz, H. Klekowicz, J. Ginter, and K. Blinowska, “On the statistical significance of event-related eeg desynchronization and synchronization in the time-frequency plane.” vol. 51, no. 7, pp. 1167–1175, July 2004.

[26] A. Delorme and S. Makeig, “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” vol. 134, no. 1, pp. 9–21, Mar. 2004.

[27] C. Bénar, M. Clerc, and T. Papadopoulo, “Adaptive time-frequency models for single-trial M/EEG analysis,” in Information Processing in Medical Imaging, ser. Lecture Notes in Computer Science, Karssemeijer and Lelieveldt, Eds., vol. 4584. Springer, 2007, pp. 458–469. [Online]. Available: http://www.springerlink.com/content/p6n9150253566121/

[28] J. Cui and W. Wong, “The adaptive chirplet transform and visual evoked potentials,” vol. 53, no. 7, pp. 1378–1384, 2006.

[29] B. Efron and R. Tibishirani, An Introduction to Bootstrap. New York: Chapman & Hall,, 1993.

[30] C. Bénar, R. Gunn, C. Grova, B. Champagne, and J. Gotman, “Statistical maps for EEG dipolar source localization,” vol. 52, no. 3, pp. 401–413, 2005.

[31] M. Gratkowski, J. Haueisen, L. Arendt-Nielsen, A. Chan, and F. Zanow, “Decomposition of biomedical signals in spatial and time-frequency modes,” vol. 47, no. 1, pp. 26–37, 2008.